

## N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM  
MICROFICHE. ALTHOUGH IT IS RECOGNIZED THAT  
CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED  
IN THE INTEREST OF MAKING AVAILABLE AS MUCH  
INFORMATION AS POSSIBLE



Technical Memorandum 80586

# Information Theory and the Earth's Density Distribution

REVISED EDITION

(NASA-TM-80586) INFORMATION THEORY AND THE  
EARTH'S DENSITY DISTRIBUTION (NASA) 43 p  
EC A03/MF A01 CSCL 08G

N80-12643

G3/46 41486  
Unclas

David Parry Rubincam

AUGUST 1979

National Aeronautics and  
Space Administration

Goddard Space Flight Center  
Greenbelt, Maryland 20771

INFORMATION THEORY AND THE EARTH'S  
DENSITY DISTRIBUTION

Revised Edition

David Parry Rubincam  
NAS-NRC Resident Research Associate  
(Presently at the Geodynamics Branch,  
Code 921, Goddard Space Flight Center)

August 1979

GODDARD SPACE FLIGHT CENTER  
Greenbelt, Maryland 20771

## CONTENTS

	<u>Page</u>
AUTHOR'S NOTE . . . . .	v
ABSTRACT. . . . .	vii
1. INTRODUCTION . . . . .	1
2. THE INFERENCE PROBLEM . . . . .	1
3. PROBABILITY . . . . .	3
4. JAYNES'S PRINCIPLE OF MINIMUM PREJUDICE . . . . .	5
4.1 Shannon's Information Measure . . . . .	5
4.2 Jaynes's Principle of Minimum Prejudice . . . . .	7
5. INFORMATION THEORY DENSITY DISTRIBUTION . . . . .	10
6. DISCUSSION . . . . .	17
6.1 Comparison with Rietsch (1977) . . . . .	17
6.2 Further Discussion of Assumptions . . . . .	19
6.3 Information Measure for Continuous Distributions . . . . .	21
6.4 Criticisms of Information Theory Inference . . . . .	28
6.5 Future Directions for the Theory . . . . .	30
ACKNOWLEDGMENTS . . . . .	31
REFERENCES . . . . .	32

## LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1 The information theory density distribution using Maxwell-Boltzmann statistics (curve A) and the opti- mum density distribution of Bullen (1975) (curve B) are shown as a function of radial distance $r$ . . . . .	37

#### AUTHOR'S NOTE

The present work is a revised version of my previous Technical Memorandum (TM), "Information Theory and the Earth's Density Distribution," NASA TM-73038, dated February 1978. There are three reasons for publishing this revision.

First, I included a discussion of Rietsch (1977). I was unaware of his important pioneering paper until I was kindly advised of it by the editor and referees of the Geophysical Journal.

Second, I expanded the discussion of certain points (such as the nature of probability) which were only briefly mentioned in the original TM.

Third and last, I presented the new material on Shannon's information measure for continuous probability distributions.

These reasons, I feel, are more than sufficient for producing a revision of the earlier work.

INFORMATION THEORY AND THE EARTH'S  
DENSITY DISTRIBUTION

David Parry Rubincam  
NAS-NRC Resident Research Associate

ABSTRACT

The "most likely" density distribution inside the earth is derived from Jaynes's (1957) information theory approach. The earth is assumed to be spherical and the density distribution spherically symmetric. The known mass and moment of inertia are used as constraints on the density distribution. The partitioning of particles among cubical boxes and use of the grand canonical ensemble from statistical mechanics result in a density distribution of the form  $\bar{\rho}(r) = 12.30 \exp(-1.46r^2/a_E^2) \text{ g/cm}^3$  where  $a_E$  is the radius of the earth. This differs from the density distribution derived by Rietsch (1977), who also used the information theory approach. The difference results from Rietsch allowing the density to vary continuously inside the volume elements rather than in discrete steps as done here. Some criticisms of information theory inference are discussed. In particular, Shannon's (1948) generalization of the information measure to continuous probability distributions is defended as the more useful measure in the continuous case over the Kullback measure. Future directions for information theory inference in solid earth geophysics are indicated.

# INFORMATION THEORY AND THE EARTH'S DENSITY DISTRIBUTION

## 1. INTRODUCTION

In a recent paper, Rietsch (1977) introduced Jaynes's (1957) information theory approach to inverse problems in solid earth geophysics. The information theory approach is a method of scientific inference which has had great success in statistical mechanics (see e.g., Jaynes 1957, 1963; Tribus 1961; Katz 1967; and Baierlein 1971) and in spectral analysis (e.g., Burg 1967, 1968, 1972; Smylie et al. 1973; and Graber 1976). Rietsch (1977) applied the approach to two problems. The first problem dealt with spectral analysis; I will not discuss it at all here. The second dealt with inferring the density distribution for the earth from knowledge of its mass and moment of inertia; the earth is assumed to be spherical and the density distribution spherically symmetric.

I have also applied information theory inference to the very same problem of the earth's density distribution. My approach, however, is somewhat different from Rietsch's (1977), and consequently so is the density distribution. In this paper I present these results together with a discussion of the differences between the two approaches, and some general comments on information theory not discussed by Rietsch (1977).

## 2. THE INFERENCE PROBLEM

This is the nature of the problem: we desire to know what the density distribution  $\rho(r)$  is as a function of radial distance  $r$  from the center of the earth, but the only information we have

about the earth is its mass  $M_E$  and moment of inertia  $C_E$ , both of which depend upon  $\rho(r)$ . Clearly we do not have enough information to say what  $\rho(r)$  actually is. Any proposed distribution which satisfies the mass and moment of inertia is nonunique; there are infinitely many other distributions which also satisfy the given data. Hence we cannot invert the data; we must infer an answer from incomplete data.

There are several methods for dealing with this problem. (For a general discussion see Bullen 1975, pp. 60-64.) The approach of Backus and Gilbert (1967, 1968) is to study all solutions consistent with the given data; this is called the geophysical inverse problem. The Backus-Gilbert approach has been used extensively. See, for example, Gilbert et al. (1973); Parker (1977a, 1977b); Jordan and Franklin (1971); and references cited by Parker (1977a, 1977b), Richards (1975), Anderson (1975), and Engdahl et al. (1975). A quite different method is that of Press (1968a, 1968b), who adopted a Monte Carlo technique of testing a wide range of models against the data and retaining only those which agreed with it. However, the commonest method by far is that of modeling: By introducing certain assumptions in addition to the data, the answer becomes unique. The assumptions of the Adams-Williamson equation and uniform chemical composition, for instance, plus the known mass, seismic velocities, and surface density determine a unique density distribution (Alterman et al. 1959, pp. 80-81). Of course a difficulty with modeling is that the assumed conditions may not hold.

Information theory inference approaches the problem from the following viewpoint: We cannot reject any possible answer (in



our case, density distribution) which agrees with the known data. We do feel, however, that some answers are more likely than others. So we do the following: Assign each possible answer a probability that it is the correct answer, then apportion the magnitudes of the probabilities in accordance with the data we have on hand. For this purpose we need to define the word "probability".

### 3. PROBABILITY

There are two major schools of thought on the nature of probability (Howard 1968, pp. 211-212). Presently, the majority of probability theory users hold that a probability is an objective quantity. A coin, for example, has a certain probability of falling head just as it has mass and angular velocity. The way to measure the probability is to flip the coin a large number of times and note the frequency of occurrence of heads.

While this is the traditional view of probability, it has the requirement of repeatability. If, for example, we discuss the probability of a successful launching of the next space station, then the objective view is of no use. The next launching is a one-of-a-kind affair, unlike the flip of a coin. The same is true of our topic: There is only one earth with one density distribution. There is no "ensemble of earths"!

This difficulty leads to the second, more powerful view of probability: Subjective probability, used in information theory inference and decision theory. The subjective view holds that a probability reflects our state of knowledge about phenomena, rather than about the phenomena themselves (Howard 1968, p. 211).

We would assign equal probabilities for a coin falling heads or tails, for instance, if we have no information which would cause us to prefer one outcome over the other. Hence a probability represents our "degree of rational belief" (Baierlein 1971, p. 13) that a particular outcome will occur. It need not be repeatable. The probabilities are subjective in the sense that they depend on a state of knowledge, and one person's data may differ from another's.

The subjective view of probability has been around for quite some time. It was held by Bayes and Laplace, and quantitative treatments have been given by John Maynard Keynes, Harold Jeffreys, John G. Kemeny, and Rudolf Carnap (Cox 1961). It was not until recently, however, that this view has gained many adherents. One reason for this situation is that it ran counter to the prevailing dogma of the objective view, as discussed by Jaynes (1967). Another reason was the lack of a cogent set of axioms from the subjective probability theorists from which to derive the probability calculus. This defect was remedied by Cox (1946, 1961), whose axioms are so simple yet compelling that they lead to the usual laws of probability without the introduction of ensembles or frequencies. This, plus the introduction of information theory by Shannon (1948), has caused the numbers of subjectivists to wax and objectivists to wane (Howard 1968, p. 212).

The probability calculus alone does not tell us how to assign probabilities; it only gives us rules for operating with them. What we need is a way of computing magnitudes of probabilities consistent with given data. This is where Jaynes's (1957)

information theory approach comes in. (Baierlein 1971 has an excellent general discussion of the information theory approach.)

#### 4. JAYNES'S PRINCIPLE OF MINIMUM PREJUDICE

##### 4.1 SHANNON'S INFORMATION MEASURE

At the heart of the approach is Shannon's (1948) information measure

$$MI(P_1, P_2, \dots, P_N) = -K \sum_{i=1}^N P_i \ln P_i \quad (4.1)$$

Here  $P_i$  is the probability that the  $i$ th of  $N$  possible answers is the correct answer and  $K$  is a positive constant. This function was originally termed the entropy function (Tribus and McIrvine 1970, p. 180), due to its similarity to thermodynamic entropy. For this reason the information theory approach is often called the Maximum Entropy Method, or MEM for short. The relationship between the information measure and thermodynamic entropy is deep, but the two are not identical (Baierlein 1971, pp. 473-478). To avoid confusion I will follow Baierlein (1971, p. 64) and call Shannon's information measure  $MI(P_1, P_2, \dots, P_N)$ , where  $MI$  stands for Missing Information, or the amount of information needed to determine which answer is correct. A better term for the approach would be ITI, or Information Theory Inference, rather than MEM.

$MI$  is not dimensionless (Edmundson, private communication, 1976), a fact that does not appear to be explicitly noted by Katz (1967) or Baierlein (1971). It carries units of information. For example, if we change the base of the logarithm in (4.1) from  $e$

to 2, which changes  $K$  to a new constant  $K'$ . and set  $K' = 1$ , then  $MI = -\sum P_i \log_2 P_i$  and  $MI$  is measured in bits. In the following development I will retain the natural logarithm base and set  $K = 1$ , so that  $MI$  is measured in nats (from natural digits; McEliece 1977, p. 15). I will also suppress the units but it should be remembered that  $MI$  is not a dimensionless quantity.

The importance of the  $MI$  function is its uniqueness; that is, given certain very reasonable assumptions of how the  $MI$  function should behave, one is inevitably led to (4.1). In this respect it is like Cox's (1946, 1961) derivation of the probability calculus.

I will not state the assumptions or prove that they lead to  $MI$ . The assumptions are given by Rietsch (1977, p. 491), and proofs are supplied by Shannon (1948, pp. 419-420) and Baierlein (1971, pp. 64-74). Rather, I will merely indicate its plausibility with an example. But first we note from (4.1) that  $MI \geq 0$ ; the amount of information needed to single out the correct answer is never negative. This is certainly an intuitively desirable property. Now let us suppose that all of the probabilities are equal. In this case it can be shown that  $MI$  attains its maximum value. This accords with intuition: we are surely in a state of maximum ignorance (i.e., need the most information) if we can favor no answer above another in terms of probability. Suppose now we have discovered that the  $j$ th possibility is the correct answer. Then  $P_j = 1$  and  $P_i = 0$  for  $i \neq j$ . How much information is missing now? In this case  $P_j \ln P_j = 1 \ln 1 = 0$ , and  $P_i \ln P_i = 0$  for  $i \neq j$  (by virtue of  $\lim_{x \rightarrow 0} x \ln x = 0$ ). Thus  $MI = 0$ ; no information is missing; we have the answer. This also accords with

intuition. Normally our ignorance lies between these two extremes, and MI takes on values accordingly between its maximum and 0. Hence MI is a plausible measure of missing information.

#### 4.2 JAYNES'S PRINCIPLE OF MINIMUM PREJUDICE

The essence of the information theory approach is this: choose the probabilities  $P_1, P_2, \dots, P_N$  of the possible outcomes to make MI as large as possible, subject to the constraints of the known data. This is Jaynes's principle of minimum prejudice (Tribus and Rossi 1973). The information theory approach is therefore rational method for assigning probabilities. In statistical mechanics, this procedure is equivalent to maximizing the entropy (Morse 1969).

To illustrate the technique with an example, suppose that we do not know the mass of the earth exactly, but do know that it must be chosen from the values  $M_1, M_2, \dots, M_N$ . Aside from  $\sum P_i = 1$ , this is all we know. We must find  $P_j$ , the probability that  $M_j$  is the correct mass, by maximizing MI. This is done by taking the partial derivative of

$$-\sum_{i=1}^N P_i \ln P_i + \alpha_0 \sum_{i=1}^N P_i$$

with respect to each  $P_i$  and setting it equal to zero. The  $\alpha_0$  is a Lagrange multiplier which ensures that all of the probabilities add up to 1. Carrying out the process yields

$$-\ln P_i - 1 + \alpha_0 = 0$$

or

$$P_i = e^{\alpha_0 - 1} = \text{constant}$$

The unknown  $\alpha_0$  may be found from the constraint

$$\sum_{i=1}^N P_i = 1$$

giving

$$P_i = 1/N$$

All of the probabilities are equal. We know nothing about the various  $M_i$  and therefore cannot favor one particular value over another.

Now suppose we obtain further information: e.g., we learn that the expectation value of the mass is

$$\sum_{i=1}^N P_i M_i = \bar{M}_E$$

We then reassign probabilities in accordance with Jaynes's principle:

$$\frac{\partial [-\sum P_i \ln P_i + \alpha_0 \sum P_i + \alpha_1 \sum P_i M_i]}{\partial P_i} = 0 \quad i = 1, 2, \dots, N$$

giving an exponential function in  $M_i$ :

$$P_i = e^{\alpha_0 - 1} e^{\alpha_1 M_i}$$

where  $\alpha_0$  and  $\alpha_1$  are Lagrange multipliers to be found from the constraints

$$\sum P_i = 1, \quad \sum P_i M_i = \bar{M}_E$$

Note that our method is completely analogous to that of the canonical ensemble in statistical mechanics (Morse 1969, pp. 268-269).

Indeed, the mathematics is identical. The only difference is in the philosophical basis, which indicates that the method has broad applicability and is not confined to statistical mechanics (Jaynes 1963, p. 192).

Maximizing the MI function is obviously the key point in the information theory approach; it provides us with the magnitudes of the probabilities. Hence justification for this approach is necessary.

The justification goes like this: MI is the unique measure for determining the amount of information needed to single out the correct answer. Any method for assigning probabilities which does not maximize MI under known constraints (knowledge) tacitly assumes information it hasn't got! In other words, if someone assigns probabilities not in accordance with Jaynes's principle, that person is prejudicing the probabilities without foundation in the known data. Thus is derived the name, "principle of minimum prejudice".

This point is particularly clear in our last example, where we knew one of the  $M_i$  was the correct answer for the mass of the earth, but had no other information (other than  $\sum P_i = 1$ ). In this case, Jaynes's principle assigned equal probabilities to all outcomes. We were completely ignorant as to which answer was correct. If someone used some other principle, and assigned, for example, a larger probability to  $M_1$  than to the other  $M_i$ , we can legitimately ask, "You favored  $M_1$  as being the most likely mass over all of the others. What basis (i.e., information) do you have for doing that?"

The user of some other principle or function also runs the risk of being inconsistent (Jaynes 1957, p. 623; Rietsch 1977, p. 493). Hence minimum prejudice, plus consistency, give the function MI a powerful claim to being the proper choice.

## 5. INFORMATION THEORY DENSITY DISTRIBUTION

I now present my own development of the information theory density distribution, and afterwards compare it to Rietsch's (1977). I will make heavy use of the methods of statistical mechanics; particularly that of the grand canonical ensemble (Morse 1969, pp. 316-327).

Imagine a three-dimensional Cartesian coordinate system with its origin at the center of the earth. The grid system will divide up the earth into many cubes of identical volumes  $V = dx \cdot dy \cdot dz$ , just as ordinary graph paper divides up a plane into squares of equal area. We can approximate the spherical surface of the earth as closely as we like by making the cubes as small as we like. Let  $\vec{r}_j$  be the vector from the center of the earth to the  $j$ th cube and set  $|\vec{r}_j| = r_j$ . Let the mass of the earth be the sum of the masses of a large number of indistinguishable particles, each with mass  $m$ . The particles are distributed amongst the cubes, with  $n_j$  particles occupying the  $j$ th cube. The mass  $M_E$  and moment of inertia  $C_E$  of the earth are then

$$\begin{aligned} M_E &= \sum_j n_j m \\ C_E &= (2/3) \sum_j n_j m r_j^2 \end{aligned} \tag{5.1}$$



where the subscript  $j$  runs over all of the cubes comprising the earth. The factor  $(2/3)$  appearing in the second equation makes use of the assumption that the density distribution is spherically symmetric, and takes care of  $r_j$  being the distance from the center of the earth to a cube and not the distance to some axis of rotation. This factor may be verified by taking the cubes to be so small that we can switch from summations to integrals without serious error, and then integrate over latitude and longitude.

Let me remark here that we have chosen cubes of equal volume so as to treat all regions of the earth identically. We have also chosen indistinguishable particles because the interchanging of particles leaves the density distribution unaffected. In other words, the only information needed to characterize the density distribution is to know the number of particles in each cube, and not to know which particular particle is in which cube. We make no commitment as to the values of  $m$  and  $V$ . As we shall see, they drop out of the final equation for the density distribution. These assumptions will be further discussed later on.

Our problem is the following. A possible model for the earth is one which has  $n_1$  particles in cube 1,  $n_2$  particles in cube 2, and so on. Each possible model will be given the subscript  $i$ , so that  $M_i$  and  $C_i$  are the mass and moment of inertia, respectively, for the  $i$ th model. We place no restrictions on the number of particles allowed to occupy each cube, so there are an infinite number of models. Our task is to assign each possible model a probability  $P_i$  that it is the correct model. Our information will

be that the expectation values of the mass  $\sum P_i M_i$  and moment of inertia  $\sum P_i C_i$  are known to be  $\bar{M}_E$  and  $\bar{C}_E$ , respectively. In practice,  $\bar{M}_E$  and  $\bar{C}_E$  are the experimentally determined values. Ultimately we will average over all of the models and find  $\bar{n}_j$ , the expectation value for the number of particles in the  $j$ th cube. From this we can determine the "most likely" density distribution.

The probabilities are computed according to Jaynes's principle of minimum prejudice:

$$\frac{\partial [-\sum P_i \ln P_i + \alpha_0 \sum P_i + \alpha_1 \sum P_i M_i + \alpha_2 \sum P_i C_i]}{\partial P_i} = 0$$

giving:

$$P_i = \frac{e^{\alpha_1 M_i + \alpha_2 C_i}}{Z}$$

where

$$Z = e^{1-\alpha_0} = \sum_i e^{\alpha_1 M_i + \alpha_2 C_i} \quad (5.2)$$

From (5.1) we can write

$$\frac{M_i}{m} = \sum_j n_{ji} \quad (5.3)$$

$$\frac{3C_i}{2m} = \sum_j n_{ji} r_j^2$$

where  $n_{ji}$  is the number of particles in the  $j$ th cube according to the  $i$ th model. The problem now looks exactly like that of the grand canonical ensemble in statistical mechanics, with  $n_{ji}$  playing the role of occupation numbers,  $r_j^2$  the energy levels, and (5.2)

the grand partition function, the equations analogous to (5.3) being

$$N = \sum_j n_j$$

$$E = \sum_j n_j \epsilon_j$$

The treatment of this problem may be found in any standard statistical mechanics text. I will follow Morse (1969, p. 326).

Using (5.3) in (5.2) we have

$$Z = \sum_i e^{\alpha_1 \sum_j n_{ji} + \alpha_2 \sum_j n_{ji} r_j^2} \quad (5.4)$$

where we have redefined  $\alpha_{1m}$  as  $\alpha_1$  and  $(2/3)\alpha_{2m}$  as  $\alpha_2$ . Note that

$$\frac{\partial \ln Z}{\partial \alpha_1} = \sum_i \left( \sum_j n_{ji} \right) \frac{e^{\alpha_1 \sum_j n_{ji} + \alpha_2 \sum_j n_{ji} r_j^2}}{Z}$$

$$= \sum_j \sum_i n_{ji} \frac{e^{\alpha_1 \sum_j n_{ji} + \alpha_2 \sum_j n_{ji} r_j^2}}{Z} = \sum_j n_{ji} P_i = \sum_j \bar{n}_j, \quad (5.5)$$

a result that we will make use of shortly.

Let us now rewrite (5.4) as a summation over the possible values of  $n_j$  instead of over  $i$ . Since there are no limits on the possible number of particles occupying each cube, we obviously have

$$Z = \sum_{n_1=0}^{\infty} e^{(\alpha_1 + \alpha_2 r_1^2) n_1} \sum_{n_2=0}^{\infty} e^{(\alpha_1 + \alpha_2 r_2^2) n_2} \dots$$

$$= Z_1 \cdot Z_2 \dots \quad (5.6)$$

where

$$z_j = \frac{1}{1 - e^{\alpha_1 + \alpha_2 r_j^2}}$$

by virtue of

$$\frac{1}{1 - x} = \sum_{n=0}^{\infty} x^n$$

so that Z separates into factors for each cube.

From (5.5) and (5.6) we have

$$\frac{\partial \ln Z}{\partial \alpha_1} = \sum_j \frac{1}{e^{-\alpha_1 - \alpha_2 r_j^2} - 1} = \sum_j \bar{n}_j$$

where evidently

$$\bar{n}_j = \frac{1}{e^{-\alpha_1 - \alpha_2 r_j^2} - 1} \quad (5.7)$$

Equation (5.7) is identical to the equation for the average number of particles in an energy state, assuming the particles follow Bose-Einstein statistics (Morse 1969, p. 326). This is hardly surprising, since the assumptions regarding the particles are the same: indistinguishability, plus no limits on the number of particles occupying a given state.

Let us now make an assumption regarding (5.7) which is well-founded in classical physics (Morse 1969, p. 325): we assume that the cubes with volume V may be taken so small that the cubes

are sparsely occupied by the particles, thus making the average number of particles in any given cube a small number compared to 1. This is equivalent to assuming the particles follow Maxwell-Boltzmann statistics (Morse 1969, p. 324), and (5.7) becomes

$$\bar{n}_j = \frac{1}{e^{-\alpha_1 - \alpha_2 r_j^2} - 1} \ll 1$$

so that

$$e^{-\alpha_1 - \alpha_2 r_j^2} \gg 1$$

and

$$\bar{n}_j \cong e^{\alpha_1} e^{\alpha_2 r_j^2}$$

The density distribution is obviously

$$\bar{\rho}(\vec{r}_j) = \frac{m}{V} e^{\alpha_1} e^{\alpha_2 r_j^2}$$

By the assumption of spherical symmetry for the density distribution we can drop the subscript  $j$  and write

$$\bar{\rho}(r) = \bar{\rho}(0) e^{\alpha_2 r^2} \quad (5.8)$$

which we take as the desired information theory density distribution. The two constants  $\bar{\rho}(0) = \frac{m}{V} e^{\alpha_1}$  and  $\alpha_2$  may be found from our

knowledge of the expectation value for the mass and moment of inertia:

$$\begin{aligned}\bar{M}_E &= 4\pi \int_0^{a_E} \bar{\rho}(r) r^2 dr = 5.976 \times 10^{27} \text{ g} \\ \bar{C}_E &= \frac{8\pi}{3} \int_0^{a_E} \bar{\rho}(r) r^4 dr = 8.068 \times 10^{44} \text{ g} \cdot \text{cm}^2\end{aligned}\quad (5.9)$$

where  $a_E$  is the radius of the earth and our numerical values have come from Stacey (1969, p. 279). We have assumed in (5.9) that the cubes are so small that we may switch from summations to integrals without serious error. By numerical integration of (5.9), or from standard mathematical tables (Graber, private communication 1978), we find that

$$\bar{\rho}(r) = 12.30e^{-1.46r^2/a_E^2} \text{ g/cm}^3 \quad (5.10)$$

is our "best guess" for the density distribution based on the given data.

A plot of (5.10) appears in Fig. 1 (refer to page 37), along with the "optimum" density distribution given by Bullen (1975, p. 361), which presumably gives the most plausible distribution on the basis of all the known data. (Rietsch 1977 also compares his curve to Bullen's.) The two curves agree remarkably well, in view of the fact that the information theory density distribution makes use of only two basic pieces of data: mass and moment of inertia. No seismic or free oscillation data have been included in our information.

## 6. DISCUSSION

### 6.1 COMPARISON WITH RIETSCH (1977)

My analysis differs from that of Rietsch (1977) at several points. Rietsch (1977) begins with Jaynes's (1968) proposed generalization of Shannon's (1948) information measure (4.1) for continuous distributions

$$MI = - \int p(\rho) \log [ p(\rho)/w(\rho) ] dv \quad (6.1)$$

where  $p(\rho)$  is the probability distribution,  $\rho$  is the density distribution, and  $dv$  is the volume element for the parameter space. The  $w(\rho)$  appearing in (6.1) is the prior probability distribution which obtains when no information is known. Equation (6.1) differs from Shannon's (1948) own proposed measure for continuous distributions in that  $w(\rho)$  does not appear. I will argue in the following paragraphs that Shannon's equation is a more useful measure than Jaynes's for continuous distributions. However, the distinction is academic in this case since Rietsch (1977) chooses a constant prior distribution  $w(\rho)$ , which for all practical purposes makes the two measures the same.

Rietsch (1977) then takes advantage of the spherical symmetry of the problem and divides up the earth with spherical shells, with the shell radii chosen so that all of the volumes between the shells are equal. Later he lets the number of shells approach infinity to obtain a continuous density distribution. I chose cubes instead of shells to lay the groundwork for the general case where there is no spherical symmetry; in particular, for finding

the density distribution which uses the known spherical harmonic coefficients of the geopotential as constraints (as explained in Section 6.5). So far the differences between approaches are minor.

The last difference of note between Rietsch's (1977) treatment and my own is the one which produces the differing equations for the density distribution. Rietsch (1977) chooses to fill the earth's volume elements with continually varying masses, rather than discrete particles. It is as though his volume elements may be filled with a continuous liquid of any amount, rather than with discrete mass-points as in my own. This leads him to a density distribution of the form (Rietsch 1977, p. 503)

$$\bar{\rho}(r) = \frac{1}{\lambda_1 + \lambda_2 r^2}$$

and from doing integrals instead of summations. His distribution has the same qualitative behavior as my own (5.10) and looks very much the same when plotted, but obviously the functional form is different. I chose to use discrete particles, since this is more in keeping with what we know about atoms, and because it more closely follows the traditional statistical mechanical development.

I should also mention that Rietsch (1977) investigated a more general distribution by putting limits on the highest and lowest density allowed in each volume element, instead of letting it vary between zero and infinity.



## 6.2 FURTHER DISCUSSION OF ASSUMPTIONS

A natural question to ask at this point is: how well do our assumptions reflect the real earth? For instance, what about using particles of equal mass? Where do atoms and molecules fit in? The answer to the last question is: they don't, at this stage. We chose this simplified model to obtain a tractable problem and illustrate the method; these are merely the first steps. If pressed upon this point, we can take our "particle" to be a proton+neutron+electron. The reason for choosing this combination is that the earth is made up predominately of elements with low atomic number. The nuclei of such elements very nearly have equal numbers of protons and neutrons. (In the heavy elements, neutrons significantly outnumber protons.) Also, electrical neutrality prevails, so that for every proton there is an electron. Hence we may think of the proton+neutron+electron as a naturally occurring unit from which the earth is made. We can then pretend that these "particles" are spread throughout the earth, and claim to know nothing of atoms, molecules, chemical bonding, etc., which would constitute further information. This line of argument also takes care of any further objection to choosing indistinguishable particles, since the previously mentioned elementary particles are indistinguishable in the fundamental sense. However, this is going to extremes.

The assumption that the cubes are sparsely occupied, thus giving Maxwell-Boltzmann statistics, may also be objectionable from an operational standpoint. After all, volumes actually

measured contain huge numbers of particles. Relaxing this condition means that we are back to Bose-Einstein statistics and (5.7) gives the average number of particles in a given cube. The density is then

$$\bar{\rho}(r) = \frac{\frac{m}{V}}{e^{-\alpha_1 - \alpha_2 r_j^2} - 1} \quad (6.2)$$

With this approach we have a problem, because there are more unknowns than constraints. If we knew how to choose  $m/V$ , then we could find  $\alpha_1$  and  $\alpha_2$  from the constraints of mass and moment of inertia, as we did before. Unfortunately, we have no clear guidance in this matter. Even if we choose  $m$  to be the mass of our "particle", we would still have to find  $V$ .

There is a way, however, to neatly sidestep the problem. We introduce a third piece of information: we assume we know  $\bar{\rho}(a_E)$ , the value of the density at the earth's surface. Using this information in (6.2) yields

$$\bar{\rho}(r) = \frac{e^{-\alpha_1 - \alpha_2 a_E^2} - 1}{e^{-\alpha_1 - \alpha_2 r^2} - 1} \bar{\rho}(a_E) \quad (6.3)$$

and we use our knowledge of  $\alpha_1$  and  $\alpha_2$  to find the two multipliers. I will not carry through the calculation, since according to Stacey (1969, p. 104) the surface density of rocks is  $2.84 \text{ g/cm}^3$ . Our Maxwell-Boltzmann equation (5.10) already gives a surface

density of  $2.86 \text{ g/cm}^3$ , so that (5.8) and (6.3) differ only trivially.

We can therefore use Bose-Einstein statistics in the information theory approach to the density distribution; but while more general than Maxwell-Boltzmann statistics, it is more complicated mathematically, as a comparison of (5.8) and (6.3) shows. The Maxwell-Boltzmann case should probably be investigated first in future developments, being simpler.

### 6.3 INFORMATION MEASURE FOR CONTINUOUS DISTRIBUTIONS

Shannon (1948, p. 628) proposed as the appropriate generalization of (4.1) for continuous distributions the function

$$MI = -K \int_{-\infty}^{\infty} p(x) \ln p(x) dx \quad (6.4)$$

where  $p(x)$  is the probability distribution and  $x$  is a continuous parameter. I will confine the discussion to the one dimensional case without loss of generality.

There are three basic objections to (6.4) being the appropriate measure (Jaynes 1963, 1968; Hobson and Cheng 1973) which are, in order of increasing seriousness: (a) It is dimensionally incorrect; (b) an infinity is thrown away in deriving it; and (c) the form of the prior probability distribution is not invariant under a change of variables. To correct these difficulties, Hobson and Cheng (1973) and Jaynes (1963, 1968) propose using the Kullback measure in its place, of which Jaynes's equation (6.1) is but a special case. However, Tribus and Rossi (1973) and Batty

(1974) argue that Shannon's original equation (6.4) is the appropriate measure.

While Rietsch (1977) follows Jaynes (1963, 1968) and Hobson and Cheng (1973), I follow Tribus and Rossi (1973) and Batty (1974), and will outline my reasons for doing so.

The problem of point (a) is the following. Suppose  $p(x) dx$  represents the probability of finding a particle in an interval  $dx$  near speed  $x$ . Then  $dx$  has dimensions cm/s and  $p(x)$  has dimensions s/cm, so that the product of the two  $p(x) dx$  is dimensionless. But the logarithm of  $p(x)$  is taken in (6.4), and this is not allowed for dimensional quantities. So (6.4) cannot be dimensionally correct.

The problem is easily remedied. As we shall see below when point (b) is discussed, the difficulty develops when  $p(x)$  is separated from  $dx$ . If we introduce a constant of value 1 and dimensions of  $x$ , then we can write

$$p(x) dx = [Dp(x)] \left[ \frac{dx}{D} \right]$$

where  $D$  is the constant. Each expression in brackets on the right side of the equation above is now dimensionless, and we can now separate  $p(x)$  from  $dx$  without falling into error with the logarithm. Since  $D$  has value 1, we can suppress it, its presence being understood. We will assume this has been done in our discussion in the following paragraphs.

As for point (b), in the continuous case  $P_i$  goes over to  $p(x) dx$  and (4.1) becomes:

$$\lim_{dx \rightarrow 0} \left\{ - \sum_i p(x_i) dx \log [p(x_i) dx] \right\}$$

$$= - \int p(x) \log [p(x)] dx - \int p(x) \log (dx) dx$$

where  $K$  has been set equal to 1. Obviously as  $dx \rightarrow 0$ ,  $\log(dx) \rightarrow -\infty$  and the right side approaches infinity. At this point the  $\log(dx)$  term is subtracted off, leaving a well-behaved function which is just (6.4). However, subtracting infinity from infinity and obtaining a finite number is usually unsound mathematically.

This can be taken care of by going back to (4.1) and writing it in exponential form

$$e^{-MI} = \prod_i \frac{P_i}{P_i}$$

Then in the continuous case it goes over to

$$e^{-MI} = \prod_i [p(x_i) dx]^{p(x_i) dx}$$

which can be written

$$e^{-MI} = \prod_i [p(x_i) p(x_i) dx] \left[ dx^{p(x_i) dx} \right] = \left[ \prod_i p(x_i) p(x_i) dx \right] [dx]$$

by noting that

$$\prod_i dx p(x_i) dx = dx \sum p(x_i) dx = dx^1$$

When more information becomes available the probability distribution  $p(x)$  goes over to some new one  $q(x)$ , whose associated measure is

$$e^{-MI_2} = \left[ \prod_i q(x_i) q(x_i) dx \right] [dx]$$

Dividing one by the other and letting  $dx$  approach 0 gives

$$\frac{e^{MI_2}}{e^{MI_1}} = \lim_{dx \rightarrow 0} \left\{ \frac{\left[ \prod_i p(x_i) p(x_i) dx \right] dx}{\left[ \prod_i q(x_i) q(x_i) dx \right] dx} \right\}$$

The  $dx$ 's cancel, giving

$$\log \left[ \frac{e^{MI_2}}{e^{MI_1}} \right] = MI_2 - MI_1 = - \int q(x) \log q(x) dx + \int p(x) \log p(x) dx$$

which is well behaved. So we may as well write (6.4) as the MI for continuous distributions, since we know now that the infinity associated with the old distribution is the same as for the new distribution, and subtraction of the two leads to no difficulties. We are therefore concerned with changes in the amount of information, and not in the amount itself.

The force of (c) may be seen from the following example. Suppose once again that  $p(x)dx$  represents the probability of finding a particle in the neighborhood  $dx$  near speed  $x$ , and further, that the speed is definitely known to lie between values  $x_1$  and  $x_2$ . We have no further information. If we apply Jaynes's principle of minimum prejudice, we find

$$\frac{\partial}{\partial p} \left[ - \int p(x) \log p(x) dx + \alpha \int p(x) dx \right] = 0$$

so that

$$-\log p(x) - 1 + \alpha = 0 ; p(x) = e^{-1+\alpha} = \frac{1}{x_2 - x_1}$$

so that  $p(x)$  is a constant for  $x_1 \leq x \leq x_2$  and zero outside the interval. But the kinetic energy  $mx^2/2$  (where  $m$  is the mass of the particle) is a perfectly respectable physical quantity. Why not take it as the continuous parameter? If we do so by setting  $y = mx^2/2$  and apply Jaynes's principle once again we find:

$$\frac{\partial}{\partial s} \left[ - \int s(y) \log s(y) dy + \alpha \int s(y) dy \right] = 0$$

which implies

$$s(y) = \frac{1}{y_2 - y_1} \quad \text{for } y_1 \leq y \leq y_2$$

The two distributions are inconsistent: a constant distribution

for speed implies a nonconstant distribution for energy and vice versa, as may be seen from

$$s(y)dy = \frac{mx \, dx}{\frac{m}{2} (x_2^2 - x_1^2)} = \frac{2x \, dx}{(x_2^2 - x_1^2)} \neq p(x)dx$$

Hence there is no one prior probability distribution. We can make it what we please for any given parameter by a suitable change of variables. To escape this difficulty, Jaynes (1968) proposes to find the prior distribution  $w(\rho)$  in (6.1) via the theory of groups. This matter is also discussed by Rietsch (1977, pp. 494-495) and by Rowlinson (1970).

I must agree with Tribus and Rossi (1973) that there really is no problem; for to change variables is to ask a different question. However, I am not entirely certain that their reasons for believing so are the same as mine, due to the terseness of their discussion. Therefore I will give my reasons below.

Let us start by recalling the meaning of (4.1). It is the amount of information needed to single out the correct answer from  $N$  possible answers. Now as  $P_i$  goes over to  $p(x) \, dx$  in the continuous case, the interpretation of (6.4) must be that it determines the amount of information needed to trap the correct answer in any one of the small intervals  $dx$ . Likewise, if we had chosen some other variable  $y = f(x)$  as the continuous parameter, then MI answers the question of how much information is needed to trap the answer in any one of the intervals  $dy$ . So when we pick  $x$  or  $y$  as the parameter, we are asking different questions



about the problem. There is no one prior distribution. The amount of information needed to answer a question is a function of the question asked.

Let us clarify the situation with our example in which  $x$  is speed and  $y$  is kinetic energy. Suppose that the mass of the particle is 2 g and the speed is definitely known to lie between 0 and 100 cm/s. Take  $dx$  to be 1 cm. The energy obviously lies between 0 and  $10^4$  ergs. Take  $dy$  to be 1 erg when  $y$  is the prior variable.

Now to use  $x$  or to use  $y$  as the parameter is to ask two different questions. Trapping the particle speed to within 1 cm/s is not the same as trapping it to within 1 erg. If  $x$  is near 50 cm/s for example, then  $dy \cong 2x dx$  and trapping the speed to within 1 cm/s means we know the energy to within  $\sim 2 \cdot 50 \cdot 1 = 100$  ergs, and not 1 erg. So it is a question of resolution. Equal resolution along the speed axis implies unequal resolution along the energy axis and vice versa. Hence it is meaningless to ask, what is the prior probability distribution? That depends on the question you are asking.

This argument is bolstered by noting an unsatisfactory aspect of (6.1) pointed out by Tribus and Rossi (1973): the information needed to single out the correct answer depends on the order in which the information is given. That it not depend on the order given is an essential part of deriving (4.1) (Baierlein 1971, pp. 64-74), and Shannon's (1948) equation is true to this condition (Tribus and Rossi 1973) when generalized to continuous distributions.

Theories are neither right nor wrong; they have only varying degrees of usefulness (Tribus 1966, p. 207). What I am suggesting here is that Shannon's (1948) original equation (6.4) is more useful than Jaynes's (1963, 1968) alternative equation when talking about information and continuous distributions.

#### 6.4 CRITICISMS OF INFORMATION THEORY INFERENCE

Some criticisms of information theory inference which have been raised will be briefly discussed here, starting with the coin flip problem.

Rowlinson (1970) argues that information theory is unable to deal with certain kinds of information. Suppose, for example, that we flip a coin 100 times and it comes up heads 75 times. Clearly we have some relevant information on whether the next flip will be heads or a tails. Rowlinson (1970) claims that information theory cannot handle this problem.

Tribus (1969), an ardent proponent of information theory inference, would probably answer this challenge according to the algorithm given on page 120 of his book: assign probabilities according to Jaynes's principle, and then modify the probabilities using Bayes's theorem when new information becomes available. In the coin flip problem the original information would be that two outcomes are possible, giving probability  $1/2$  to heads and tails. The new information would be the 75 heads out of 100 flips. This would be used in Bayes's theorem to give the new probabilities. I will not pursue this problem further here.

Another apparent drawback is that information theory inference at times gives "unphysical" answers. For instance, if we omitted our knowledge of the moment of inertia in solving for the density distribution, then we would obtain a constant density all throughout the earth, as may be easily verified. This does not seem reasonable; we feel that the density should certainly increase towards the center of the planet. The value of information theory inference appears questionable in this instance.

The problem is easily resolved, as may be seen in the following example. Suppose that instead of guessing the earth's density distribution we were confronted with a small object of exotic shape and unknown composition and asked to guess its density distribution on the basis of known mass and volume. In this case a constant density distribution does not appear at all unreasonable; this is because we are in a state of extreme ignorance about the object. With the earth, however, this is not the case: we have some ideas about how the earth ought to behave. In this instance it is that gravity should pull the heavier material towards the center of the earth, and high interior pressures will compress it, making the density increase towards the center. Hence we are dealing with tacit information. We can hardly withhold information from the method and then criticize it for not reproducing what we did not tell it! So if an answer appears "unphysical", then we have not been fair to the method; we did not tell it everything we knew.

## 6.5 FUTURE DIRECTIONS FOR THE THEORY

The results obtained above may be easily generalized to include any known volume integrals of the density distribution.

Supposing that there are L such integrals having the form

$$\int_{\substack{\text{volume} \\ \text{of earth}}} \rho(\vec{r}) f_i(\vec{r}) dv = F_i \quad (i = 1, \dots, L)$$

the resulting average density distribution is

$$\bar{\rho}(\vec{r}) = \text{const} \cdot \exp(\alpha_1 f_1(\vec{r}) + \alpha_2 f_2(\vec{r}) + \dots + \alpha_L f_L(\vec{r})) \quad (6.5)$$

The Lagrange multipliers  $\alpha_i$  are to be found from the known values  $F_i$ . Note that the above result is not restricted to the spherically symmetric case. Besides the mass and moment of inertia, the spherical harmonic coefficients  $\bar{C}_{\ell m}$  and  $\bar{S}_{\ell m}$  of the earth's gravitational field immediately come to mind as integrals having this form. I intend to publish the resulting  $\bar{\rho}(\vec{r})$  based on the gravity field coefficients in the near future.

The next obvious extension of the theory is to assume that the earth is an elastic body so as to include the elastic parameters  $\mu(\vec{r})$  and  $\lambda(\vec{r})$  in addition to the density distribution  $\rho(\vec{r})$  as unknown quantities to estimate. This will allow seismic travel times, free oscillation periods, and body tide observations to be used, all of which depend upon  $\mu(\vec{r})$ ,  $\lambda(\vec{r})$  and  $\rho(\vec{r})$ . Graber (1977) has already made a start in this direction using mass, moment of inertia, and three zero-node torsional normal modes of degree  $\ell = 2, 8$ , and 26 of the earth. More realistic treatment of atoms and molecules has already been mentioned. Information theory

inference should be compared to other inverse techniques, such as the Backus-Gilbert method. Gull and Daniell (1978) briefly discuss the two methods. The goal of information theory inference is to put in all of the physics and data we know about the earth and maximize the remaining missing information.

Since there will never come a day when we have all of the information, solid earth geophysics will always have a need for sound methods of inference. Information theory is such a method. Its philosophical basis is satisfying: no unwarranted weighting of possible answers. It is rational and objective: Everyone using it will obtain the same answers, given the same data (once the formulation of the problem is agreed upon!); it gives the "best" answer on the basis of very little data; it provides an alternative to extensive modeling; and its mathematics is standard—that of statistical mechanics. Information theory inference should find extensive use in solid earth geophysics.

#### ACKNOWLEDGMENTS

I wish to thank Michael A. Graber and David E. Smith for many helpful discussions. I thank Professor H. P. Edmundson of the University of Maryland for pointing out that MI is not dimensionless. The support of a NAS-NRC Resident Research Associateship is gratefully acknowledged.

## REFERENCES

- Alterman, Z., H. Jarosch and C. L. Pekeris, 1959. Oscillations of the earth, Proc. Roy. Soc. London, 252A, 80-95.
- Anderson, D. L., 1975. Structure of the mantle, Rev. Geophys. Space Phys., 13, No. 3, 300-302.
- Backus, G. E. and J. F. Gilbert, 1967. Numerical applications of a formalism for geophysical inverse problems, Geophys. J. R. Astr. Soc., 13, 247-276.
- Backus, G. E. and J. F. Gilbert, 1968. The resolving power of gross earth data, Geophys. J. R. Astr. Soc., 16, 169-205.
- Baierlein, R., 1971. Atoms and Information Theory, W. H. Freeman, San Francisco.
- Batty, M., 1974. A comment on the paper "A comparison of the Shannon and Kullback information measures," J. Stat. Phys., 11, 523-524.
- Bullen, K. E., 1975. The Earth's Density, Chapman and Young, London.
- Burg, J. P., 1967. Maximum entropy and spectral analysis, paper presented at the 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, OK.
- Burg, J. P., 1968. A new analysis for time series data, paper presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, Netherlands.

- Burg, J. P., 1972. The relationship between maximum entropy spectra and maximum likelihood spectra, Geophysics, 37, 375-376.
- Cox, R. T., 1946. Probability, frequency, and reasonable expectation, Am. J. Phys., 14, 1-13.
- Cox, R. T., 1961. The Algebra of Probable Inference, Johns Hopkins Press, Baltimore.
- Engdahl, E. R., A. M. Dziewonski, and E. A. Flinn, 1975. Structure of the earth's core, Rev. Geophys. Space Phys., 13, no. 3, 302-303.
- Gilbert, J. F., A. M. Dziewonski, and J. Brune, 1973. An informative solution to a seismological inverse problem, Proc. Nat. Acad. Sci. USA, 70, 1410-1413.
- Graber, M. A., 1976. Polar motion based upon Doppler, IPMS and BIH data, Geophys. J. Roy. Astr. Soc., 46, 75-85.
- Graber, M. A., 1977. An information theory approach to the density distribution of the earth. NASA TM 78034, November 1977.
- Gull, S. F., and G. J. Daniell, 1978. Image reconstruction from incomplete and noisy data, Nature, 272, 686-690.
- Hobson, A., and B. K. Cheng, 1973. A comparison of the Shannon and Kullback information measures, J. Stat. Phys., 7, 301-310.
- Howard, R. A., 1968. The foundations of decision analysis, IEEE Trans. Systems Science and Cybernetics, SSC-4, 211-219.

Jaynes, E. T., 1957. Information theory and statistical mechanics, Phys. Rev., Series 2, 106, 620-630, and 108, 171-190.

Jaynes, E. T., 1963. Information theory and statistical mechanics, in Statistical Physics, pp. 181-218, K. W. Ford, editor, W. A. Benjamin, New York.

Jaynes, E. T., 1967. Foundations of probability theory and statistical mechanics, in Delaware Seminar in the Foundations of Physics, pp. 77-101, M. Bunge, editor, Springer-Verlag, New York.

Jaynes, E. T., 1968. Prior probabilities, IEEE Trans. Systems Science and Cybernetics, SSC-4, 227-241.

Jordan, T. H., and J. N. Franklin, 1971. Optimal solutions to a linear inverse problem in geophysics, Proc. Nat. Acad. Sci. USA, 68, 291-293.

Katz, A., 1967. Principles of Statistical Mechanics, W. H. Freeman, San Francisco.

McEliece, R. J., 1977. The Theory of Information and Coding, Addison-Wesley, Reading, Mass.

Morse, P. M., 1969. Thermal Physics, second edition, W. A. Benjamin, New York.

Parker, R. L., 1977a. Linear inference and underparameterized models, Rev. Geophys. Space Phys., 15, 446-456.



Parker, R. L., 1977b. Understanding inverse theory, Ann. Rev. Earth Planet. Sci., 5, 35-64.

Press, F., 1968a. Earth models obtained by Monte Carlo inversion, J. Geophys. Rev., 73, 5223-5234.

Press, F., 1968b. Density distribution in earth, Science, 160, 1218-1221.

Richards, P. G., 1975. Theoretical seismology, Rev. Geophys. Space Phys., 13, No. 3, 295-298.

Rietsch, E., 1977. The maximum entropy approach to inverse problems, J. Geophys., 42, 489-506.

Rowlinson, J. S., 1970. Probability, information and entropy, Nature, 225, 1196-1198.

Shannon, C. E., 1948. A mathematical theory of communication, Bell System Tech. J., 27, 379-423, and 623-656. (Also, reprinted in The Mathematical Theory of Communication by C. E. Shannon and W. W. Weaver, University of Illinois Press, Urbana, IL 1949.)

Smylie, D. E., G. K. Clarke, and T. J. Ulrych, 1973. Analysis of irregularities in the earth's rotation, in Methods of Computational Physics, 13, pp. 391-430, B. A. Bolt, editor, Academic Press, New York.

Stacey, F. D., 1969. Physics of the Earth, first edition, Wiley, New York.

Tribus, M., 1961. Thermostatics and Thermodynamics, Van Nostrand, New York.

Tribus, M., 1966. Micro- and macro-thermodynamics, American Scientist, 54, 201-210.

Tribus, M., 1969. Rational Descriptions, Decisions, and Designs, Pergamon, New York.

Tribus, M., and E. C. McIrvine, 1971. Energy and information, Sci. Am., 225, no. 3, 179-188.

Tribus, M., and R. Rossi, 1973. On the Kullback information measure as a basis for information theory: comments on a proposal by Hobson and Chang, J. Stat. Phys., 9, 331-338.

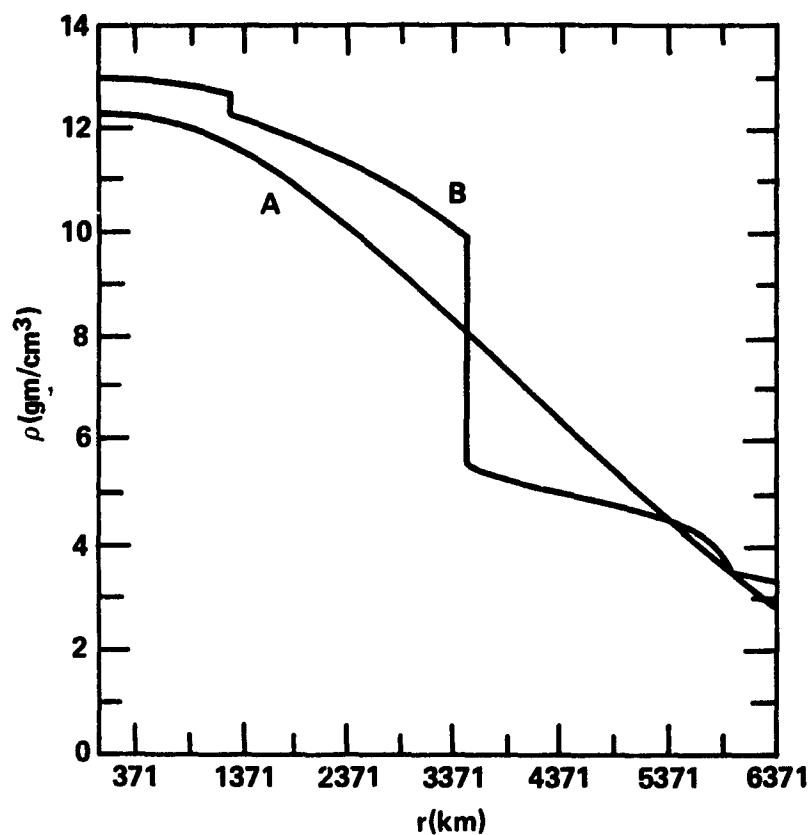


Figure 1. The information theory density distribution using Maxwell-Boltzmann statistics (curve A) and the optimum density distribution of Bullen (1975) (curve B) are shown as a function of radial distance  $r$ .

# BIBLIOGRAPHIC DATA SHEET

1. Report No. TM 80549	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Information Theory and the Earth's Density Distribution, Revised Edition		5. Report Date August 1979	
		6. Performing Organization Code 921	
7. Author(s) David Parry Rubineam		8. Performing Organization Report No.	
9. Performing Organization Name and Address Goddard Space Flight Center Geodynamics Branch, Code 921 Greenbelt, MD 20771		10. Work Unit No.	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address  Same as above		13. Type of Report and Period Covered	
		14. Sponsoring Agency Code	
15. Supplementary Notes  Submitted to Geophysical Journal			
16. Abstract  The present paper argues for using the information theory approach of Jaynes(1957) as an inference technique in solid earth geophysics. A spherically symmetric density distribution is derived as an example of the method. A simple model of the earth plus knowledge of its mass and moment of inertia leads to a density distribution which is surprisingly close to the optimum distribution of Bullen(1975). Future directions for the information theory approach in solid earth geophysics as well as its strengths and weaknesses are discussed.			
17. Key Words (Selected by Author(s)) Information theory, density distribution		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 37	22. Price*

\*For sale by the National Technical Information Service, Springfield, Virginia 22151.

GSFC 25-44 (10/77)